UNITED STATES PATENT APPLICATION

FOR

METHOD AND APPARATUS FOR MEASURING USER ACCESS TO IMAGE DATA

Inventors: Simon S. Chen
              Yen Whei Chow
              Todd Tao Zhou

1

2 ## BACKGROUND OF THE INVENTION

3

4 1.   FIELD OF THE INVENTION

5   The present invention relates to the field of network analysis in general, and in

6 particular, to HTTP based network analysis.

7

8 2.   DESCRIPTION OF THE RELATED ART

9   Many, if not most of Internet based businesses depend on advertising for

10 revenue generation.   One common method of generating revenue is to charge for

11 displaying the advertisements or banner images of third parties.   In some cases,

12 instead of charging fees, or as partial consideration for displaying such ad banner

13 images, an exchange program is arranged whereby two entities agree to display each

14 other's banner images on their respective Internet sites.   As with any form of

15 advertising, it is important to know how many persons are viewing the particular

16 advertisements or banner images, and what percentage of viewers respond to

17 advertisements by clicking on the ads or by responding to the ads in some measurable

18 manner.

19   In the sense that revenue is often advertising based, Internet-based business

20 opportunities can be equated to the television industry.   In the television industry, the

21 Nielsen™ rating system is perhaps one of the best known media measurement

22 systems.   Established in the 1950's, the Nielsen rating system today utilizes

23 monitoring devices at a set of selected user sites to monitor television viewing habits.

1   The Nielsen rating system generates statistical information regarding the number of

2   viewers who have viewed programming on a particular television channel during a

3   particular period.

4   The Nielsen rating system does not provide information regarding the

5   advertisements that were watched by the viewers. For example, the Nielsen rating

6   system may report that 10 million viewers watched a particular television episode

7   during one particular week. However, no indication is provided regarding the number

8   of viewers that watched a particular advertisement -- which was shown during that

9   television episode and was also shown at other times, on the same and other channels

10   -- during that week.

11   A system other than the above-described program rating system collects data

12   on advertisements which are broadcast. It does this by essentially monitoring all

13   television channels and collecting data on the number of times a particular

14   advertisement is broadcast. This system monitors the source of the advertisement (by

15   monitoring the television broadcasts) and, therefore, cannot directly provide

16   information on the number of viewers who viewed a particular advertising campaign

17   during a particular time period. While this data may be combined with data from the

18   Nielsen rating system in order to estimate the number of times a particular

19   advertisement was viewed, this process is, of course, cumbersome and not always

20   accurate.

21   Further, and perhaps of more relevance to the present invention, it is

22   essentially not possible to collect data from all "broadcasts" at the source in a

1   distributed network such as the Internet -- simply because there are too many (perhaps

2   hundreds of thousands, if not millions) of sources of advertisements.

3   Any number of Internet statistics gathering tools have become available in

4   recent years. In general, these tools can be divided into two categories. First, a large

5   number of tools are available for gathering statistics at the source, e.g., the individual

6   servers. These tools can provide information on the number of Internet pages served,

7   the number of advertisements served, etc. Unfortunately, because they are gathering

8   information from the individual sources, these tools cannot provide a complete

9   picture of the penetration of a full advertising campaign and they are limited in ability

10   to provide information on the demographics of the individuals viewing the

11   advertisements.

12   Tools are also available to gather information at the viewer's site.

13   Unfortunately, these tools are also limited in their information gathering capability.

14   For example, it is often reported that a particular number of viewers viewed a

15   particular uniform resource locator (URL) during a particular time period.

16   Unfortunately, these tools are not able to report information on individual

17   advertisements viewed. For example, even if it is known that a URL identifies an

18   advertisement, the URL does not necessarily uniquely identify any particular

19   advertisement. This is in part because the advertisements are often "served" from an

20   ad server which rotates advertisement banner image images under the same URL.

21   What is needed is a system which can accurately measure the number of on-

22   line users that are presented with specific advertisements, and which can provide

1   additional statistical reporting regarding user interaction with specific advertisements

2   or other image data.

3   Accordingly, it is an object of the present invention to provide a method and

4   apparatus which accurately measures the number of times a banner image image (or

5   other image) is viewed by a network user, and which identifies the unique images

6   viewed by each particular on-line user.

7   It is still another object of the present invention to accomplish the above-stated

8   objects by utilizing a method and apparatus which is simple in use and design, and

9   efficient in reducing interference with the normal operation of a user's computer.

10   The foregoing objects and advantages of the invention are illustrative of those

11   which can be achieved by the present invention and are not intended to be exhaustive

12   or limiting of the possible advantages which can be realized. Thus, these and other

13   objects and advantages of the invention will be apparent from the description herein

14   or can be learned from practicing the invention, both as embodied herein or as

15   modified in view of any variation which may be apparent to those skilled in the art.

16   Accordingly, the present invention resides in the novel methods, arrangements,

17   combinations and improvements herein shown and described.

1 ## SUMMARY OF THE INVENTION

2

3   In accordance with these and other objects of the invention, a brief summary

4 of the present invention is presented. Some simplifications and omissions may be

5 made in the following summary, which is intended to highlight and introduce some

6 aspects of the present invention, but not to limit its scope. Detailed descriptions of a

7 preferred exemplary embodiment adequate to allow those of ordinary skill in the art to

8 make and use the inventive concepts will follow in later sections.

9   According to broad aspects of the invention, methods and apparatuses for

10 providing information regarding the number of visits to pages on a data network such

11 as the Internet and banner images encountered on network pages are described. The

12 described embodiments overcome a number of issues faced by prior art systems,

13 including providing for improved accuracy in measuring the number of times a banner

14 image or advertisement is viewed; providing improved methods and apparatuses for

15 efficiently identifying unique banner images viewed; providing an improved method

16 and apparatus for configuring a network user's computer so that interference from the

17 collection of data with the normal operation of the computer is minimized; providing

18 an improved method and apparatus for efficiently calculating an image checksum to

19 allow unique identification of a banner image viewed by an end user; and providing

20 an improved method and apparatus for determining whether the network user has

21 used the *BACK* button of an Internet browser to view a page and, if so, to accurately

22 count the number of banner images viewed.

23

1

2 <u>**BRIEF DESCRIPTION OF THE DRAWINGS**</u>

3

4   Figure 1 is a representation of an Internet page as may be monitored by an

5 embodiment of the present invention.

6   Figure 2 is an overall diagram of a network as may be utilized by an

7 embodiment of the present invention.

8   Figure 3A is a high level block diagram of a first embodiment of a client

9 computer as may be utilized by the present invention.

10   Figure 3B is a high level block diagram of a second embodiment of a client

11 computer as may be utilized by the present invention.

12   Figure 4 is a flow diagram illustrating a data collection method as may be

13 implemented by an embodiment of the present invention.

14   Figure 5 is a flow diagram illustrating a method of identifying banner images

15 in Internet pages as may be utilized by the present invention.

16   Figure 6 is a representation of an Internet page using frames as may be

17 monitored by an embodiment of the present invention.

18   Figure 7 is a flow diagram illustrating a method of monitoring frame pages as

19 may be utilized by an embodiment of the present invention.

20   Figure 8 is a flow diagram illustrating a method of BACK button processing

21 as may be utilized by an embodiment of the present invention.

22   Figure 9 is a diagram illustrating certain panel member demographics which

23 may be utilized by an embodiment of the present invention.

1   Figure 10 is an illustration of a report format as may be utilized by an

2   embodiment of the present invention.

3   Figure 11 is an overall flow diagram of a method of retrieving images as may

4   be utilized by the present invention.

5   For ease of reference, the numerals in all of the accompanying drawings are

6   usually in the form "drawing number" followed by two digits, xx; for example,

7   reference numerals on Figure 1 may be numbered 1xx; on Figure 3, reference

8   numerals may be numbered 3xx. In certain cases, a reference numeral may be

9   introduced on one drawing and the same reference numeral may be utilized on other

10   drawings to refer to the same item.

11

1   <u>**DETAILED DESCRIPTION OF**</u>

2   <u>**THE EMBODIMENTS THE PRESENT INVENTION**</u>

3

4

5   *I.*   *<u>OVERVIEW OF HTML FOR BANNER IMAGES</u>*

6   Figure 1 illustrates an Internet page 101 which includes a separate image 102

7 that could be a hyperlink represented as a graphic "button", or a banner containing an

8 advertisement. The image 102 is also referred to herein as a "banner image," "image,"

9 "advertisement" "banner" or simply an "ad." A network user viewing the Internet

10 page (a "viewer," "end user" or "panel member") may ignore the banner image 102,

11 simply look at the banner image 102 or, more actively, select the banner image 102

12 (such as by clicking on it with a cursor control device). By selecting the banner image

13 102, the viewer may be presented with another Internet page which may provide, for

14 example, another page of information or another page providing more detail on a

15 company placing an advertisement or on a product being advertised in the banner

16 image 102. Alternatively, the banner image 102 may provide one form or another of

17 rich new media such as audio or video programming content.

18   Internet pages are typically constructed using a programming language called

19 hypertext markup language (HTML). It is, in fact, the HTML code which is

20 transmitted from an Internet server to the requesting machine in response to a viewer

21 requesting a particular Internet page or site (identified by its uniform resource locator

22 or "URL"). Internet pages which include banner images 102 have encoded in their

23 HTML what will be termed herein "anchor pairs". An anchor pair comprises the

24 HTML code for the URL to contact if the user selects the banner image 102, together

1   with the URL for the image to display in the banner. An example of an anchor pair is

2   shown below in Table I.

3
4
5
6
7
8
9
10

TABLE I
ANCHOR PAIR
href="http://www.digitalriver.com/dr/v2/ec_MAIN.Entry17c? CID=5560&SID=6505&SP=10007&PN=5&PID=100853">Buy Speedlane Software Online!</A> </FONT></B></P><TABLE WIDTH="120" BORDER="0" CELLPADDING="0" CELLSPACING="0" ALIGN="RIGHT"><TR> <TD><IMG SRC="/graphics/spacer.gif" WIDTH="20" HEIGHT="4" BORDER="0" ALIGN="BOTTOM"></TD><TD><a

11
12

13   There is not necessarily a one-to-one correspondence between advertising

14   images and the URL encoded in the HTML for the anchor pair. In fact, there may be

15   a many-to-many correspondence. For example, the advertising image may be

16   provided from an advertising server. Thus, the particular image served may vary

17   every time that an Internet page is accessed although the URL for the page remains

18   constant. An example of the HTML for this is shown in Table II.

19

20
21
22
23
24
25
26
27
28

TABLE II
ANCHOR PAIR
   <img src="http://209.1.112.252/adgraph/follow.gif" width=468 height=60 alt="[Click our Sponsor's banner, with Easy Return to Hotmail.]" hspace=0 vspace=0 border=0></td></tr>

1

2

3   Moreover, the same advertising image may be associated with any number of

4   URLs. For example, a particular advertiser may contract with multiple advertising

5   server companies to place its advertisement on multiple Internet pages. There will be

6   at least one, if not many, different URLs used by each advertising server company to

7   serve the advertisement.

8   Thus, it is not possible to accurately track the number of times an

9   advertisement is viewed by simply tracking URLs.

10

11   *II.*   *OVERVIEW OF AN EXEMPLARY EMBODIMENT FOR*

12   *TRACKING INTERNET BASED ADVERTISMENT VIEWING*

13   Similar to the Nielsen rating system, it is possible to recruit a panel of viewers

14   which provide a statistically representative sample of a population of data network

15   users, such as Internet users, in order to provide statistically interesting data regarding

16   data access habits and preferences.

17   In one exemplary embodiment, an index group of approximately 2000 Internet

18   users was developed using random digit dialing to insure demographic accuracy and

19   projectability of the panel member's behavior to the population of Internet users.

20   After demographic profiles of the index panel were established, an additional 23,000

21   (for 25,000 total) members that fit the demographic profiles were selected via Internet

22   recruiting. Internet recruiting is a relatively cost effective method of recruiting panel

23   members. Periodic, e.g., quarterly, re-calibration of the index panel is employed in

1   the process of recruiting new panel members to reflect the changing population of the

2   Internet user community.

3   When a panel member is selected, the panel member completes a survey

4   which identifies certain key demographic and psychographic data to allow a profile of

5   the user to be built. As will be described below, the panel member then instructs his

6   or her computer to allow the collection of information regarding advertisements

7   received by the panel member's computer while the panel member is "surfing the

8   Internet".

9

10   *III.*   *OVERALL ARCHITECTURE*

11   Figure 2 provides a high level overall view of the architecture of one preferred

12   embodiment of the present invention. In Figure 2, the general relationship among the

13   features of the system is shown as used in a distributed network environment 210

14   such as the Internet.

15   A plurality of panel member client/viewer terminal devices or computers 201

16   are configured to collect information relating to specific banner images 102, such as

17   advertisements. These advertisements are typically viewed as a result of accessing

18   world wide web sites or pages on the Internet 210. The panel member computers 201

19   may be based on any of a number of platforms executing various operating systems

20   and browsers. For example, the platform may be executing any of a number of

21   different operating systems including UNIX, the Macintosh OS™, or the Windows™

22   operating system. The platform may also be executing any of a number of Internet

23   browsers including, for example, browsers available from Netscape Corporation or

1   Microsoft Corporation or browsers available from online service providers such as

2   AOL, Compuserve or Prodigy. Advantageously, the present invention requires little,

3   if any, modification for use on these varying platforms and is relatively simple to

4   install.

5   It should be understood that the references to specific programs or components

6   typically found in general purpose computer terminals and servers, related to but not

7   forming part of the invention, are provided for illustrative purposes only. References

8   to computer programs and components are provided for ease in understanding how

9   the present invention may be practiced in conjunction with known types of on-line

10   database and data network/Internet applications. Moreover, it is important to

11   understand that the various components of the system contemplated by the present

12   invention may be implemented by software programs, by direct electrical connection

13   through customized integrated circuits, or a combination of circuitry and

14   programming, using any of the methods known in the industry for providing the

15   functions described herein without departing from the teachings of the invention.

16   Those skilled in the art will appreciate that from the disclosure of the invention

17   provided herein, both programming languages and commercial semiconductor

18   integrated circuit technology would suggest numerous alternatives for actual

19   implementation of the functions herein that would still be within the scope of the

20   present invention.

21   In one preferred embodiment, the computers 201 are further configured with a

22   proxy server architecture. Use of the proxy server architecture provides a number of

1 advantages including ease of portability from platform to platform. The proxy server

2 architecture will be described in greater detail with reference to Figures 3A & 3B.

3 Data is collected by a proxy server 306 when a panel member's computer 201

4 accesses a distributed network 210. The collected data is transmitted back over the

5 distributed network 210, in this example the Internet, and is reported to a panel server

6 221. The collected data includes, among other items, a banner image link URL, a

7 banner image URL, and a checksum/length field for each banner image 102 presented

8 to or viewed by a panel member. The panel server 221 receives the collected data,

9 and logs it in one or more data logs 307.

10 The panel server 221 preferably executes on a NT/Pentium based general

11 purpose computer. In the described embodiment, a plurality of panel servers 221 are

12 provided in order to assure high availability and fast user access. The particular

13 number of panel servers 221 may vary from embodiment to embodiment and may

14 depend on such as factors as the size and speed of the panel server 221, the number of

15 panel members in the sample population, etc.

16 The panel server 221 also provides the collected data to a database server 233

17 for further processing. The database server 233 performs the function of overall

18 database management for the system of the present invention. In the described

19 embodiment, an Oracle relational database server is utilized. However, alternative

20 embodiments may utilize any of a number of database servers and, in fact, the

21 database server 233 may utilize either a relational or non-relational database without

22 departure from the spirit and scope of the present invention.

1   In the described embodiment, there are two main sources of data. First,

2   demographic data is collected and stored with respect to the makeup of the members

3   of a panel. The demographic data may include information such as gender, age,

4   marital status, educational level, race, employment status, income level, industry of

5   employment, occupation, and geographic region information. It is anticipated that a

6   panel of 25,000 members will generate about 300MB of data per day, to be received

7   and processed by the database server 233.

8   The database server 233 stores the banner images 102 for each unique banner

9   image 102 that is encountered. The database server 233 performs the function of

10   correlating the foregoing data to generate reports, as will be described in greater detail

11   below.

12   Periodically (e.g., daily), an analysis engine 234 analyzes the data correlated

13   by the database server 233 and stored in the database. The analysis engine 234

14   performs several tasks, including that of obtaining the banner images 102 for each

15   advertisement presented to a panel member. As described above, there is a many-to-

16   many relationship between the advertisement images and the URLs. A method for

17   determining the particular advertisement image viewed is described in greater detail

18   below.

19   Subscribers to the system may access the database in order to obtain reporting

20   on advertisements viewed. In the described embodiment, the subscribers may access

21   the database through a HTTP server 235. In alternative embodiments, subscribers

22   may be given alternative access. For example, subscribers may be given direct dial-in

23   access or may be provided with reports periodically by facsimile, mail or email.

1

2

3

4   *IV.*   *CONFIGURATION OF THE PANEL MEMBER'S COMPUTER*

5   One method of configuring a panel member's computer is illustrated generally

6   in an exemplary embodiment shown in Figure 3A. In Figure 3A, a panel member's

7   computer 201 is configured by installing metering software 303 designed to intercept

8   messages communicated between the operating system 304 and a browser 305. While

9   this technique may be utilized in certain embodiments of the present invention, design

10   and development of metering software 303 for each of the many platforms which may

11   need to be supported is likely to be cumbersome because the metering software 303

12   must be customized for each browser/operating system combination. It should be

13   noted that configuration of a panel member's computer 201 may be accomplished by

14   any of a number of techniques that implement the foregoing functions without

15   departing from the inventive aspects of the present invention. For example, in the

16   embodiment described above, the present invention combines the proxy server 306

17   with a browser 305 to intercept messages communicated between the operating

18   system 304 and a browser 305 (see Figure 3B).

19   It has been discovered that it is advantageous to configure the

20   computer 201 as illustrated in Figure 3B, by providing the proxy server 306 to collect

21   data related to the banner images 102 accessed by a panel member. One distinct

22   advantage of use of the proxy server 306 over metering software 303 is that use of the

23   proxy server 221 allows for the development of relatively portable code.

1   *V.*   *SYSTEM OPERATION*

2

3   The components of Figure 3B are best understood by referring to the system's

4   data collection process illustrated in the flowchart shown in Figure 4. In operation, a

5   panel member first selects a URL using any of a number of conventional browsing

6   methods, such as selecting a hyperlink or directly typing the URL into the an Internet

7   browser 305 (Block 401). The proxy server 306 intercepts the URL request (Block

8   402) and passes the URL request onto the Internet 210, where the request is served in

9   the conventional manner (Block 403).

10   The proxy server 306 then initiates generation of what will be termed a

11   "captured data record" (Block 404). The captured data record provides information

12   relating to the URL request, the HTML data received, the panel member's use of the

13   Internet page, and advertising banner images 102 encountered on the Internet page. In

14   one embodiment of the present invention, the captured data record preferably

15   comprises the information identified below in Table III:

16

17   TABLE III

	FIELD	DESCRIPTION
1	VERSION NUMBER	Version number of proxy software
2	SITE ID	Used by the panel server and database server to identify the panel member's computer
3	USER ID	Used by the panel server and database server to identify the panel member
	REQUESTED URL	The URL requested by the panel member

4		
5	METHOD	HTTP methods supported by the target of the hypertext link. The most common methods are GET, HEAD and POST.
6	REFERRER	The URL of the referring page (only applicable in the case of a hyperlink)
7	REQUEST TIME OF URL (GMT)	The time of day that the user requested the URL (in GMT)
8	REQUEST TIME OF URL (LOCAL)	The time of day that the user requested the URL (in local time)

1

2   In addition, the following fields, shown in Table IV are generated or collected

3 for each banner image 102 found in the HTML page that is viewed:

4

	FIELD	DESCRIPTION
9	BANNER IMAGE ANCHOR URL	The URL of the banner image 102 anchor (page to go to if the panelist clicks on the banner image 102)
10	BANNER IMAGE URL	The URL of the banner image 102
11	CHECKSUM	A calculated checksum for the banner image 102.
12	LENGTH	The length of the banner image 102 in bytes

5

6   The length of each captured data record is approximately 500 bytes. Keeping

7 the amount of captured data which must be transmitted to the panel server 221

8 minimal is important to avoid undue interference with the performance of the panel

9 member's computer 201. The operation of the present invention must be as

10 unobtrusive as possible so that it does not unnecessarily interfere with the panel

11 member's experience while accessing the Internet. Interference with the panel

12 member's experience may result in changes in the behavior of the panel member and,

1   in the case of significant interference, may result in the panel member removing

2   himself or herself from the pool of panel members.

3   It should be noted that in alternative embodiments, alternative types of

4   browsing data may be transmitted with the captured data record, which may have an

5   impact on the overall length of the captured data record and the level of useful

6   information collected. For example, in addition to transmitting the URL of the banner

7   image 102, the full image may be transmitted. While transmitting the full banner

8   image 102 may provide useful information for the analysis engine 234, transmission

9   of the full banner image 102 is relatively expensive both in terms of bandwidth

10   consumed in transmission of the image and in terms of storage requirements.

11   Instead of transmitting the data for each entire banner image 102, a checksum

12   is preferably calculated for the banner image 102 and reported in the captured data

13   record. In one embodiment of the present invention, the checksum is calculated

14   against only a sampling of the banner image 102. The amount of image data sampling

15   is variable, and can be set based on the desired exactness in identifying specific

16   banner images 102. By calculating the checksum against only a sampling of the

17   banner image 102, processing bandwidth is saved when compared with calculating

18   the checksum for the entire image. For example, in the described embodiment, only

19   recurrent bytes (e.g., every 4th or 5th byte) are used in the checksum calculation.

20   While using only a portion of the banner image 102 to calculate a checksum

21   can advantageously reduce processing requirements, it does not provide the same

22   level of assurance that the checksum will represent a unique value identifying, for

23   example, an advertisement, as would be provided if the checksum were calculated for

1   the entire banner image 102. As can be understood, varying the checksum sampling

2   rate allows for varying the reliability of the results against the benefit of saving

3   computational cycles and bandwidth.

4   At times there may be only minute differences between two images 102, such

5   as where two advertisements are produced by a single advertiser. In such a case, if

6   the differences do not occur in the recurrent bytes sampled to generate the checksum,

7   the checksum will not uniquely identify the advertisement image. To overcome this

8   problem, the total length of the advertising image is calculated in addition to the

9   checksum. In one embodiment of the present invention, the length of the banner

10   image 102 in bytes is determined and provided in the captured data record for the

11   page.

12   This combination of checksum and length values are used to uniquely identify

13   each specific banner image 102 that is encountered. It is been determined empirically

14   that, while not providing absolute assurance that the checksum/length combination

15   will always identify a specific advertising image, the use of the combined

16   checksum/length value is sufficiently reliable for purposes of the described

17   embodiment.

18   It is worthwhile pointing out that in alternative embodiments, alternative

19   information may be used to uniquely identify a banner image 102. One example was

20   briefly discussed above—storing and transmitting the entire banner image 102, with

21   the inherent sacrifice in storage and transmission bandwidth. As also discussed

22   above, a checksum could be calculated on the entire banner image 102 with the

23   inherent additional costs in processing, storage and transmission requirements. For

1 purposes of the discussion herein, data uniquely identifying a banner image 102,

2 regardless of the method used to generate the identifying information, will be referred

3 to generically as a "unique banner image identifier". Generating a unique banner

4 image identifier for identifying a specific image eases the process of counting and

5 analyzing the number of times a particular image has been displayed.

6   Unlike the banner image data, certain of the fields in the captured data record

7 may be determined prior to receiving the HTML data (e.g., USER ID and REQUEST

8 TIME OF URL) while other fields will necessarily have to be determined after the

9 HTML data is received. In any event, the HTML data corresponding to the requested

10 URL is eventually received by the proxy server 306 (Block 405). The proxy server

11 306 then passes the HTML data onto the browser 305 (Block 406).

12   As one important aspect of the present invention, the proxy server 306

13 examines the HTML data to find additional banner images 102. Each captured data

14 record may include data relating to $0-n$ banner images 102, depending on the number

15 of banner images 102 found in the HTML data. The proxy server 306 completes its

16 generation of the captured data record and communicates the captured data record

17 over the network 210 to data log 307 (Block 407). The data are also communicated

18 over the network 210 to the panel server 221 (Block 408).

19   Turning now to Figure 5, a method of identifying banner images 102 as may

20 be implemented in the described embodiment is illustrated. Initially, the HTML code

21 of a page that a panel member is viewing is scanned for anchor/banner image 102

22 pairs (Block 501). As described above, anchor/banner image 102 pairs contain the

1　HTML code for the URL to contact if the user selects the banner image 102, together

2　with the URL for the image to display in the banner 102.

3　　　The system of the present invention scans the entire HTML page for all

4　anchor/banner image 102 pairs, and if no anchor/banner image 102 pair is found, then

5　the process completes without going through any banner identification (Block 503 to

6　END).

7　　　If a pair of anchor/banner images 102 is found (Block 503), the present

8　invention (optionally) filters the anchor/banner image 102 pairs to screen out images

9　which do not likely represent banner images 102 based on the image size (Block 504).

10　For example, images such as graphic "buttons" to be clicked on for hyperlinking

11　could be confused for advertisements if any image size is accepted. Image size is

12　determined by multiplying the width of the image times the height of the image (in

13　pixels). One embodiment of the present invention uses a minimum image size

14　threshold to filter images. In another embodiment, the filtering process requires that

15　the image size exceed a first threshold but be smaller than a second threshold.

16　　　The filter thresholds in the described embodiment are variable, and may be set

17　based on empirical observations that the size of particular banner images 102, such as

18　advertisements, likely fall within a certain range. For example, as the size of

19　advertising banner images 102 becomes increasing standardized, it should be easier to

20　filter out images which do not fit within one of the standard sizes.

21　　　If an image does not pass the filtering process (Block 506), the system then

22　checks if more HTML code is present and reverts to Block 501 to continue scanning

23　the remainder of the HTML code for any banner images 102 that may be present.

1   After all of the HTML code is scanned and no images are found, the process is

2   completed. If an image does pass through the preset thresholds of the filtering

3   process (Block 506), then the combination checksum/length value is computed for the

4   banner image 102 in the process described above to identify the specific

5   advertisement (Block 508). The entire process is completed for each image found as

6   the remainder of the HTML code of the page is scanned (Block 509).

7   The system of the present invention is designed to perform the foregoing

8   processes even if the HTML page received utilizes frames technology. An HTML

9   page using frames is shown in Figure 6.   Since there are 3 sub-pages in the

10   exemplary page illustrated by Figure 6, there will be 4 URLs downloaded by the

11   browser. They are represented generally as:

12

13   http://domain.com/mainframe.html
14   http://domain.com/sub-page1.html
15   http://domain.com/sub-page2.html
16   http://domain.com/sub-page3.html
17

18   The downloading sequence is typically the "Main frame" first, followed by the

19   three sub-pages. The three sub-pages are downloaded concurrently via multithreads

20   by the browser 305. As was described above, the proxy server 306 is designed to

21   transmit to the panel server 221 one captured data record for each HTML page

22   viewed. In non-frames HTML, a single HTML page corresponds to a single URL

23   being downloaded by the proxy server 306. As is seen, in a frame HTML page, a

24   single page may require multiple URL requests. However, it is still desirable to send

25   a single data record that corresponds to the panel member's access of the multi-frame

1   page. Thus, as another aspect of the present invention, a method is disclosed for

2   detecting that a HTML page is a frame page and transmitting a single captured data

3   record to the panel server 221 for each frame page.

4   Referring now to Figure 7, the method is described in greater detail. Initially,

5   each page of HTML code that is received is parsed to identify the HTML tag

6   "FRAME" or "IFRAME" (Block 701). If the tag is not found (Block 702), the page is

7   identified as not being a main page for a frame, and is processed (searching for banner

8   images 102, adding up the page length, etc.) in accordance with the methods

9   described above (Block 703).

10   If the tag is found, the system initiates the identification of any sub-frames that

11   may exist. As understood by those skilled in the art, sub-pages of a frame are

12   typically received by the user's computer 201 within a predetermined amount of time

13   after the main frame is received. In the present invention, all pages received before

14   the next hyperlink selection or the entering of a URL by a panel member (a page with

15   a FRAME tag), are identified as sub-pages (Block 704). The length of all sub-pages

16   is included with the length determined for the main page, and the combination of data

17   is included in the captured data record for the main page (Block 705). In addition, all

18   banner images 102 in each of the sub-pages is identified using the processes described

19   above, and the data for such images 102 are generated along with the captured data

20   record of the main page (Block 706). As can be seen, the data related to each sub-

21   page is handled in combination with the data for the main page of a multi-frame page.

22   Turning now to Figure 8, a method for accounting for use of the BACK button

23   of a browser 305 is explained. When a user clicks the BACK button of the browser

1 program (Block 801), the browser 305 usually displays a page from its cache memory.

2 If the page is retrieved from cache, it may not be reported by the proxy server 306 and

3 thus, an inaccurate count of the number of times a particular Internet page (and the

4 associated advertisements or banner images 102) is viewed will result. Thus, as one

5 aspect of the described embodiment, the proxy server 306 forces a reload of the

6 HTML code every time that the user selects the BACK button in order to accurately

7 calculate the number of times a banner image 102 is actually viewed.

8   The reloaded page normally has HTTP status code 304: no new content

9 (Block 802). Thus, if a page has banner images 102 and the reload page is returned

10 with a status code 304, special handling of the HTML page is provided in the present

11 invention in order to avoid the loss of banner image 102 information. This handling

12 is done in one of two ways dependent on whether the banner image 102 is static or

13 dynamic.

14   *Static banner images* -- Static banner images are banner images 102 which do

15 not change each time a browser reloads a HTML page. Therefore, when the user

16 selects the BACK button, the static banner images 102 in that re-visited page do not

17 change and the user sees the same banner image 102 again. As was just mentioned,

18 when the HTML page has a status code 304, there is no new content and therefore the

19 proxy server 306 does not parse the HTML code for banner images 102. According

20 to one aspect of the present invention, when the proxy server 306 detects the status

21 code 304, it sends a message to the panel server 221 stating that the previous page has

22 already been visited (Block 803). The panel server 221 communicates the message to

23 the database server 233. The analysis engine 234, which is configured to recurrently

1   search its records, will check for the previously visited page (by matching URLs) and

2   copy the banner image 102 information associated with the previously visited page

3   into a new data capture record (Block 804).

4   Assume, for example, the user visits an Internet page http://domain.com/

5   page1.html with 2 banner images B1 and B2. The proxy server 306 will send a

6   message to the panel server 221 with the content: http://domain.com/ page1.html,

7   200, B1, B2, where 200 is the status code for the page (normal). If the user then visits

8   another page, http://domain.com/page2.html, the proxy server 306 sends a message

9   with the content: http://domain.com/page2.html, 200. If the user then selects the

10   BACK button of the browser 305, the record: http://domain.com/ page1.html, 304 is

11   sent to the panel server 221, inserted into the database server 233 and then the

12   analysis engine 234 searches its previous records for the entries for the page

13   http://domain. com/page1.html and copies the banner images 102 from that entry such

14   that the final entry in the database server 233 records is:

15   http://domain.com/page1.html, 304, B1, B2.

16   It should be noted that in an alternative embodiment, the records for

17   previously visited pages may be stored and searched locally at the client system. This

18   would, however, add overhead processing to the client system.

19   *Dynamic banner images* -- Dynamic banner images are banner images 102

20   which change each time a page is accessed even if the HTML page which contains the

21   banner images 102 does not change. It is possible that an Internet page contains both

22   static and dynamic banner images 102. For example, assume page1 contains two

23   banner images 102 (as was described in the previous example), banner images B1 and

1   B2. Assume that banner image B1 is a static banner image 102 and banner image B2

2   is a dynamic banner image 102. When the user selects the BACK button of the

3   browser 305, the user sees a different banner image 102 (banner image 102 B3) in

4   place of banner image 102 B2.

5   The present invention will record the fact that banner image 102 B1 and B3

6   were viewed when the BACK button was selected. As discussed above, a

7   checksum/length value is calculated for each banner image 102 that is viewed. In the

8   example given above, the first time that the user visited the Internet page, the

9   length/checksum was calculated for banner images B1 and B2 as:

10   B1, L1, C1
11   B2, L2, C2
12   *(where Bn=banner/anchor pair; Ln=banner length; Cn=checksum)*
13
14   This length and checksum information will be sent to the panel server 221 as

15   part of the data capture record for the HTML page.

16   According to the BACK button process of one embodiment of the present

17   invention, the second time the user visits the page by selecting the BACK button, the

18   HTML page is returned with a no new content status having a status code 304 (Block

19   801 & 802). The dynamic banner image 102 uses the same URL as the original

20   banner image 102, however its content is changed. An image (for banner image 102

21   B3) is received by the panel member's computer 201 (Block 812). The banner image

22   102 information (e.g., B3, L3, C3) is sent to the panel server 221 indicating that the

23   HTML page was revisited, along with an image summary for the new image B3

24   (Block 813). The panel server 221 then updates the data capture record by searching

1   its database, replacing the data related to the first dynamic banner image 102 with the

2   data related to the new banner B3 (Block 814).

3   As has been discussed, one of the difficulties in collecting and analyzing

4   information regarding advertisements or banner images 102 on the Internet is that

5   there is a many-to-many relationship between the advertisements and URLs

6   identifying the advertisements. It has now been described that for each

7   advertisements viewed, the panel member's computer 201 reports, among other data,

8   the banner image URL, a banner image checksum and a banner image length. The

9   analysis engine 234 uses this information to uniquely identify the advertisements

10   viewed.

11   Turning to Figure 9, an overall flow diagram for finding an actual banner

12   image 102 viewed by a panel member is shown. As has been described, for each

13   HTML page viewed by a panel member, information collected and prepared in a data

14   capture record is sent from the panel member's computer 201 to a proxy server 306

15   and eventually, to database server 233 for analysis by analysis engine 234. The

16   information contained in a data capture record, detailed in Tables III and IV, includes

17   for each banner image 102, the banner image 102 anchor URL, the banner image 102

18   URL, the banner image 102 checksum and the banner image 102 length (as shown in

19   Table IV).

20   The first time a banner image 102 is accessed by a panel member's computer

21   201, the banner image 102 is stored in the database 223. Stored banner images 102

22   are also referred to as "banner image masters". A banner image master comprises the

23   image together with the checksum/length calculated for the image. Each time a

1   banner image 102 is encountered while a user is browsing the Internet, the checksum

2   and length of the a banner image 102 are compared with the checksum/length

3   combinations for previously accessed banner images 102 stored in the database

4   (Block 901). If a match is found (branch 903), the stored banner image 102 is

5   assumed to be the image viewed (Block 904). The data related to the new banner

6   image 102 is not stored in the database, rather the image data is discarded.

7   If the checksum/length of the new banner image 102 is not found in the

8   database (branch 906), the distributed network (Internet) 210 is then accessed at the

9   indicated URL of the new banner image 102 (Block 912) and the checksum/length is

10   again computed for the retrieved banner image 102 (Block 913). The

11   checksum/length value is computed again because the banner image 102 may, for

12   example, be retrieved from an advertising server. Thus, many ads may match the

13   particular URL, but the checksum/length value for the retrieved banner image 102

14   may or may not match the checksum/length value for the banner image 102 viewed.

15   If there is not a match (branch 915), the distributed network 210 is accessed again to

16   obtain a different banner image 102, and the process of computing the

17   checksum/length value and comparing it to those values in the database is repeated

18   until a pre-selected retry limit is exceeded (branch 919).

19   In some cases, the particular image 102 may not be available from the

20   advertisement server and, as a result, no matter how many times the process is

21   repeated the image will not be found. Thus, a retry limit is imposed. If the retry limit

22   is exceed (branch 920), an entry is made in the database indicating that a banner

1   image 102 having a checksum/length value matching the reported checksum length

2   was not found in the distributed network 210 (Block 921).

3   If a match was found during one of the retry processes (branch 916), the image

4   and its checksum/length value are added to the database (Block 922).

5   Table V further illustrates the processing performed by the analysis engine 234

6   for possible HTML return codes and banner image 102 information (see Table III and

7   IV), the cause associated with the return codes, and the processing required by the

8   analysis engine 234 for handling particular page conditions. In Table V, "An"

9   represents the anchor link of banner image 102, "In" represents the image of the

10   banner image 102, "Ln" represents the image length, "Cn" represents the image

11   checksum, "-1" for the length represents an unknown image length and Ax,Ix,Lx,Cx

12   represents any other existing data.

13

14   TABLE V

15   HTML RETURN CODE / BANNER IMAGE 102

16   INFORMATION PROCESSING

Case	Why It Happens	Process Needed
200 only	Full HTML page retrieved, page contains no banner image 102	Normal process; send information from Table III to panel server

200+An+In+Ln+Cn	Full HTML page retrieved, page contains banner images(s) 102	1. If (An,In) does not exist, new banner image 102 master will be created with (Ln,Cn) 2. If (An,In) exists with (-1,0), replace this banner image 102 with data (Ln,Cn) 3. If (An,In) exists with multiple (Ln,Cn), create a new one.
200+An+In+-1+0	Full HTML page retrieved. Page contains banner image 102(s) but the banner image 102 is already in browser's cache.	1. If (An,In) does not exist, new banner image 102 master should be created with (-1,0). 2. If (An,In) exists and only has one instance of (Ln,Cn), do not create new banner image master. Existing banner image 102 will be used. 3. If (An,In) exists with multiple (Ln,Cn), random pick one.
304 only	HTML page in cache. No image(s) is loaded by browser.	1. Copy all banner images 102 from latest 200 page. 2. If no 200 page is found, ignore banner images 102.

| 304+An+In+Ln+Cn | 1. HTML page in cache.
2. New banner image 102 found. Banner image 102(s) can be created from sub-frame page or Java script.
3. Image 102 is retrieved also. | 1. Copy banner images 102 from latest 200 page.
2. If (An,In,Ln,Cn) exists, ignore the new banner image 102.
3. If (An,In)s exist but have different (Lx,Cx), replace all copied (An,In,Lx,Cx) with new (An,In,Ln,Cn).
4. If (An)s exist but have different (Ix,Lx,Cx), replace all copied (An,Ix,Lx,Cx) with (An,In,Ln,Cn).
5. If no match, create one.
Note: All (An,In,Ln,Cn) etc. in 304 case only talk about the banner image 102 instances copied from 200 page. |
| 304+An+In+-1+0 | 1. HTML in cache.
2. New banner image 102 found.
3. Banner image 102 is in browser's cache, so no banner image 102 is reloaded. | 1. Copy banner images 102 from latest 200 page.
2. If (An,In) exists, use copy version
3. If (An) exists, replace (An,Ix,Lx,Cx) with (An,In,-1,0)
4. If no match and there is only one banner image 102 in 200 page, drop old one use new one (An,In,-1,0)
5. If no match and there are multiple banner images 102 in 200 page, create a new banner image 102. |

304+null+In+Ln+C n	1. HTML page in cache 2. New image(s) is retrieved	1. Copy banner images 102 from latest 200 page 2. If (Ax,In,Lx,Cx) exists, replace it with (Ax,In,Ln,Cn) 3. If no match, ignore
304+null+In+-1+0	1. HTML page in cache 2. Image reloaded but either the image is redirected to a cached image or returned with 304	ignore

VI. *SUBSCRIBER REPORTING*

4 Once the foregoing data has been collected, the system of the present

5 invention generates comprehensive subscriber reports. The reports include data

6 detailing top Internet sites accessed during a particular period, Internet site reports

7 detailing specific information on activity at particular sites, and ad summary reports

8 summarizing information relating to particular advertisements or banner images 102.

9 The reports may cover any given time period, for example, weekly, monthly or

10 quarterly time period.

11 In particular, in the described embodiment, five reports are provided showing

12 information relating to top Internet sites including: (i) Top Internet Sites by Unique

13 Site, (ii) Top Internet Sites by Property, (iii) Top Referring Sites by Unique Site, (iv)

14 Top Internet Sites by Domain and (v) Top Navigation Guides by Unique Site. The

15 reports provide information regarding site audience, Internet activity and profile

1   information which include rank, unique audience size, reach, page views, pages

2   viewed from browser cache and pages viewed per person. The SITE_ID and

3   USER_ID are used to uniquely identify a user profile in order to provide demographic

4   information for reporting.

5   In addition to these reports, on-line access to the database is provided by, for

6   example, the HTTP server 235 (see Figure 2) which allows template-driven queries,

7   thereby providing customized reports. Other reports available include (i) a

8   Demographic Targeting--Site report providing statistically significant sites based on

9   selected audience characteristics; (ii) a Demographic Targeting--Banner Image report

10   which provides data related to the statistically significant banner images 102 viewed

11   by the target audience; (iii) an Audience Profile--Site report which profiles and

12   compares up to three selected sites demographics, unique audience, composition and

13   coverage site; (iv) an Audience Profiles --Banner Image report which provides

14   audience profiles for selected banner images 102 and includes unique audience,

15   composition, impressions, click rate, reach and frequency with all demographic

16   groupings.

17   What has been described herein is a method and apparatus for accurately and

18   efficiently counting the number of times an image 102 is viewed by a user of an on-

19   line database or data network, such as the Internet. Although the present invention

20   has been described in detail with particular reference to preferred embodiments

21   thereof, it should be understood that the invention is capable of other and different

22   embodiments, and its details are capable of modifications in various obvious respects.

23   As is readily apparent to those skilled in the art, variations and modifications can be

1   affected while remaining within the spirit and scope of the invention. Accordingly,

2   the foregoing disclosure, description, and figures are for illustrative purposes only,

3   and do not in any way limit the invention, which is defined only by the claims.

4

5